



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### Lower glycolysis carries a higher flux than any biochemically possible alternative

**Citation for published version:**

Court, SJ, Waclaw, B & Allen, RJ 2015, 'Lower glycolysis carries a higher flux than any biochemically possible alternative', *Nature Communications*, vol. 6, 8427. <https://doi.org/10.1038/ncomms9427>

**Digital Object Identifier (DOI):**

[10.1038/ncomms9427](https://doi.org/10.1038/ncomms9427)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Publisher's PDF, also known as Version of record

**Published In:**

Nature Communications

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## ARTICLE

Received 27 Mar 2014 | Accepted 21 Aug 2015 | Published 29 Sep 2015

DOI: 10.1038/ncomms9427

OPEN

# Lower glycolysis carries a higher flux than any biochemically possible alternative

Steven J. Court<sup>1</sup>, Bartłomiej Waclaw<sup>1</sup> & Rosalind J. Allen<sup>1</sup>

The universality of many pathways of core metabolism suggests a strong role for evolutionary selection, but it remains unclear whether existing pathways have been selected from a large or small set of biochemical possibilities. To address this question, we construct *in silico* all possible biochemically feasible alternatives to the trunk pathway of glycolysis and gluconeogenesis, one of the most highly conserved pathways in metabolism. We show that, even though a large number of alternative pathways exist, the alternatives carry lower flux than the real pathway under typical physiological conditions. We also find that if physiological conditions were different, different pathways could outperform those found in nature. Together, our results demonstrate how thermodynamic and biophysical constraints restrict the biochemical alternatives that are open to evolution, and suggest that the existing trunk pathway of glycolysis and gluconeogenesis may represent a maximal flux solution.

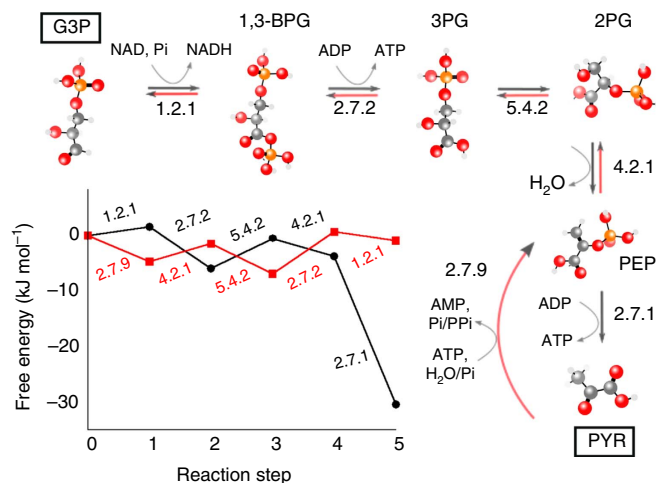
<sup>1</sup>SUPA, School of Physics and Astronomy, University of Edinburgh, James Clerk Maxwell Building, Peter Guthrie Tait Road, Edinburgh EH9 3FD, UK. Correspondence and requests for materials should be addressed to R.J.A. (email: rallen2@staffmail.ed.ac.uk).

The biochemical pathways of central carbon metabolism are highly conserved across all domains of life, and largely control the productivity of life on the Earth<sup>1,2</sup>. Yet it remains unknown whether these pathways are the result of historical contingency during early evolution<sup>3,4</sup>, or are instead optimal solutions to the problem of energy and biomass production<sup>5–10</sup>. Put simply, are there alternative biochemically feasible pathways that could perform the same function, and if so, how do they perform compared with those found in nature?

Previous studies have mainly addressed this question either by constructing simplified artificial metabolic networks<sup>5–7</sup>, or by mining databases of biochemical compounds and reactions for known organisms<sup>4,8,9,11–13</sup>. Both of these approaches have drawbacks: the former does not capture real biochemistry, whereas the latter is limited to metabolites and reactions found in well-studied organisms. Nevertheless, this work has led to suggested optimality principles including maximizing biochemical flux or yield<sup>6,7</sup>, minimizing biochemical steps or protein costs<sup>5,8</sup> and ensuring that function is maintained in a changing environment<sup>11</sup>. Only a few studies have attempted to explore the full universe of possible metabolic pathways, using realistic rules of biochemistry<sup>10,14–17</sup>. In particular, using this approach, Noor *et al.*<sup>10</sup> recently suggested that central carbon metabolism, taken as a whole, uses the minimal number of enzymatic steps needed to generate a predefined set of biochemical precursors.

Here, we perform an exhaustive computational search for all possible biochemically feasible alternatives to one of the most ancient and most highly conserved sections of central carbon metabolism, the trunk pathway of glycolysis and gluconeogenesis. Glycolysis breaks down glucose to pyruvate, generating ATP, NADH and biosynthetic precursors, whereas gluconeogenesis uses ATP and NADH to generate glucose from pyruvate. The glycolytic-gluconeogenic pathway is almost linear and can be divided into two parts: an ‘upper’ chain of reactions involving 6-carbon molecules, which connects glucose to glyceraldehyde-3-phosphate (G3P), and a ‘lower’ chain of reactions involving 3-carbon molecules, which connects G3P to pyruvate (Fig. 1). This lower reaction chain is known as the trunk pathway. In prokaryotes, the glycolytic and gluconeogenic trunk pathways consist of almost the same set of five reactions, differing only in one step. In glycolysis, the exergonic conversion of phosphoenolpyruvate (PEP) to pyruvate is coupled to the phosphorylation of ADP to ATP, whereas in gluconeogenesis the reverse reaction is driven by the hydrolysis of ATP to AMP, either with release of inorganic phosphate (the phosphoenolpyruvate synthase (pps) route) or via consumption of inorganic phosphate and release of pyrophosphate (the phosphate dikinase (ppdk) route).

Although the upper part of the glycolytic/gluconeogenic pathway exists in several very distinct variants, most notably the Embden–Meyerhof–Parnas (EMP) and Entner–Doudoroff (ED) pathways (Supplementary Fig. 1)<sup>8,18</sup>, the trunk pathway is near-ubiquitous, and contains enzymes which are highly conserved and universally distributed across the three domains of life<sup>19,20</sup>. Only a very small number of variations are observed in nature (see Supplementary Discussion for details). The existence of such an ancient and universal pathway suggests three possible scenarios: (i) the trunk pathway is the only biochemical possibility, (ii) alternatives exist but the extant pathway is evolutionarily optimal and (iii) alternatives are possible but have not been found by evolution. Distinguishing these scenarios lies at the heart of much of evolutionary biology. Here, we address this question directly using a computational approach. By systematically constructing and exploring the full space of biochemically feasible metabolites and reactions, many of which are not currently exploited by any



**Figure 1 | The trunk pathway of glycolysis and gluconeogenesis.** The end points are glyceraldehyde 3-phosphate (G3P) and pyruvate (PYR); intermediate metabolites are 1,3-bisphosphoglycerate (1,3-BPG); 3-phosphoglycerate (3-PG); 2-phosphoglycerate (2-PG) and phosphoenolpyruvate (PEP). For each reaction, the external metabolites involved and the first three numbers from the EC number classification are indicated. The inset shows thermodynamic profiles for the trunk pathway in the glycolytic and gluconeogenic (pps) directions (see Methods for details).

known organism, we find that hundreds of alternative trunk pathways are possible. The one observed in nature, however, carries the maximal biochemical flux, under reasonable constraints on the intermediate metabolite and enzyme concentrations. Our results suggest that the trunk pathway represents an optimal solution among many possible alternatives.

## Results

**A network of all possible biochemical reactions.** We used a computer program to generate the network of all possible biochemically feasible pathways between G3P and pyruvate.

**Metabolites.** Our program first systematically generates all possible relevant internal metabolites, that is, molecules intermediate between G3P and pyruvate, including those that are not found in nature (see Methods for details). Central carbon metabolism consists exclusively of reactions between ‘CHOPN’ molecules: those composed of carbon, hydrogen, oxygen, phosphorus and nitrogen atoms, with phosphorus being present only in phosphate groups<sup>21,22</sup>. Moreover, the trunk pathway contains only unbranched aliphatic 3-carbon CHOPN molecules. For completeness, we therefore include in our analysis all possible unbranched aliphatic CHOPN molecules containing up to four carbon atoms. We consider only molecules that are negatively charged (that is, include carboxyl or phosphate groups). This condition is motivated by the need to avoid leakage through the negatively charged phospho-lipid membrane, and is satisfied by almost all molecules in core metabolism<sup>23</sup>. Applying these criteria results in almost 1,500 different molecules, including all the internal metabolites in the real trunk pathway. We computed the free energy of formation  $\Delta_f G$  for all our internal metabolites, using existing experimental data where possible<sup>24</sup> or, in the absence of such data, using a variant of the group contribution method<sup>25–28</sup> (see Methods for details).

**Reactions.** Our program generates all possible reactions among our set of internal metabolites, based on the 20 Enzyme Commission (EC)

**Table 1 | The set of enzymatic reaction types used in generating the network of reactions.**

EC class		Example reaction
1.1.1	Oxidation	$\text{CH}_3\text{-CH}_2(\text{OH}) + \text{NAD}^+ \rightleftharpoons \text{CH}_3\text{-CHO} + \text{NADH}$
1.2.1	phosphorylation	$\text{CH}_3\text{-CHO} + \text{NAD}^+ + \text{P}_i \rightleftharpoons \text{CH}_3\text{-COP} + \text{NADH}$
1.3.1		$\text{COOH-CH}_2\text{-CH}_2\text{-COOH} + \text{NAD}^+ \rightleftharpoons \text{COOH-CH=CH-COOH} + \text{NADH}$
1.4.1	Deamination	$\text{COOH-CH(NH}_2\text{)-CH}_3 + \text{NAD}^+ + \text{H}_2\text{O} \rightleftharpoons \text{COOH-CO-CH}_3 + \text{NADH} + \text{NH}_3$
2.6.1	Transamination	$\text{COOH-CH(NH}_2\text{)-CH}_3 + 2\text{-oxoglutarate} \rightleftharpoons \text{COOH-CO-CH}_3 + \text{glutamate}$
2.7.1	Phosphate transfer	$\text{CHO-CH}_2(\text{OH}) + \text{ATP} \rightleftharpoons \text{CHO-CH}_2\text{P} + \text{ADP}$
2.7.2		$\text{CH}_3\text{-COOH} + \text{ATP} \rightleftharpoons \text{CH}_3\text{-COP} + \text{ADP}$
2.7.9		$\text{CH}_3\text{-CO-COOH} + \text{ATP} + \text{H}_2\text{O/P}_i \rightleftharpoons \text{CH}_2 = \text{Cp-COOH} + \text{AMP} + \text{P}_i/\text{PP}_i$
3.1.3	Hydrolysis	$\text{CH}_3\text{-CH}_2\text{P} + \text{H}_2\text{O} \rightleftharpoons \text{CH}_3\text{-CH}_2(\text{OH}) + \text{P}_i$
3.5.1		$\text{CH}_3\text{-CO(NH}_2\text{)} + \text{H}_2\text{O} \rightleftharpoons \text{CH}_3\text{-COOH} + \text{NH}_3$
3.6.1		$\text{CH}_3\text{-COP} + \text{H}_2\text{O} \rightleftharpoons \text{CH}_3\text{-COOH} + \text{P}_i$
4.1.1	Decarboxylation	$\text{CH}_3\text{-CH(OH)-COOH} + \text{H}_2\text{O} \rightleftharpoons \text{CH}_3\text{-CH}_2(\text{OH}) + \text{CO}_2(\text{aq})$
4.2.1	Dehydration	$\text{CH}_3\text{-CH(OH)-COOH} \rightleftharpoons \text{CH}_2 = \text{CH-COOH} + \text{H}_2\text{O}$
4.3.1	Ammonia-lyase	$\text{COOH-CH(NH}_2\text{)-CH}_3 \rightleftharpoons \text{COOH-CH=CH}_2 + \text{NH}_3$
5.3.1	Isomerization	$\text{CH}_3\text{-CO-CH}_2(\text{OH}) \rightleftharpoons \text{CH}_3\text{-CH(OH)-CHO}$
5.3.2		$\text{COOH-CH}_2\text{-CO-COOH} \rightleftharpoons \text{COOH-CH=C(OH)-COOH}$
5.4.2		$\text{CH}_3\text{-CH(OH)-CH}_2\text{P} \rightleftharpoons \text{CH}_3\text{-CHp-CH}_2(\text{OH})$
5.4.3		$\text{COOH-CH(NH}_2\text{)-CH}_2\text{-CH}_3 \rightleftharpoons \text{COOH-CH}_2\text{-CH(NH}_2\text{)-CH}_3$
6.3.1	ATP-driven amine ligase	$\text{R-COOH} + \text{ATP} + \text{NH}_3 \rightleftharpoons \text{R-CO(NH}_2\text{)} + \text{ADP} + \text{P}_i$
6.4.1	ATP-driven carboxylation	$\text{CH}_3\text{-CH}_2(\text{OH}) + \text{CO}_2(\text{aq}) + \text{ATP} \rightleftharpoons \text{CH}_3\text{-CH(OH)-COOH} + \text{ADP} + \text{P}_i$

The set of reaction types included in our analysis, defined by the first three numbers of the EC classification. Although we show non-ionic formulae, all compounds are assumed to exist as an equilibrium mixture of protonated and deprotonated species in solution (Supplementary Methods). Phosphate groups  $\text{-PO}_4^-$  are denoted by a 'p'. Not all variants of each reaction class are listed here; for a complete list, see Supplementary Table 3.

reaction classes<sup>29</sup>, which encompass all the reactions between CHOPN molecules in core metabolism, of length four carbon atoms or less (Table 1). Many of these reactions involve cofactors such as ATP and  $\text{NAD}^+$ ; we refer to these as external metabolites and we assume that they have fixed concentrations, which define the cellular (physiological) conditions. The full set of external metabolites in our reaction network is shown in Supplementary Table 1 and includes the energy currency metabolite ATP, the redox cofactor  $\text{NAD}^+$ , the amino-group donor glutamate, as well as inorganic phosphates and  $\text{CO}_2$ .

**Our network generates many alternative trunk pathways.** Our network reveals a huge number of alternative pathways connecting G3P and pyruvate, which are consistent with the rules of biochemistry. Enumerating these alternatives, we find 2, 29, 555 and 5,859 pathways of length 3, 4, 5 and 6, respectively, with the number of pathways increasing exponentially with path length (Supplementary Fig. 2). Some of these alternative pathways are found to use the same set of reaction types as the natural trunk pathway, but execute them in a different order (for example, left pathway of Fig. 2a), whereas others make use of a different set of reaction types (right pathway of Fig. 2a).

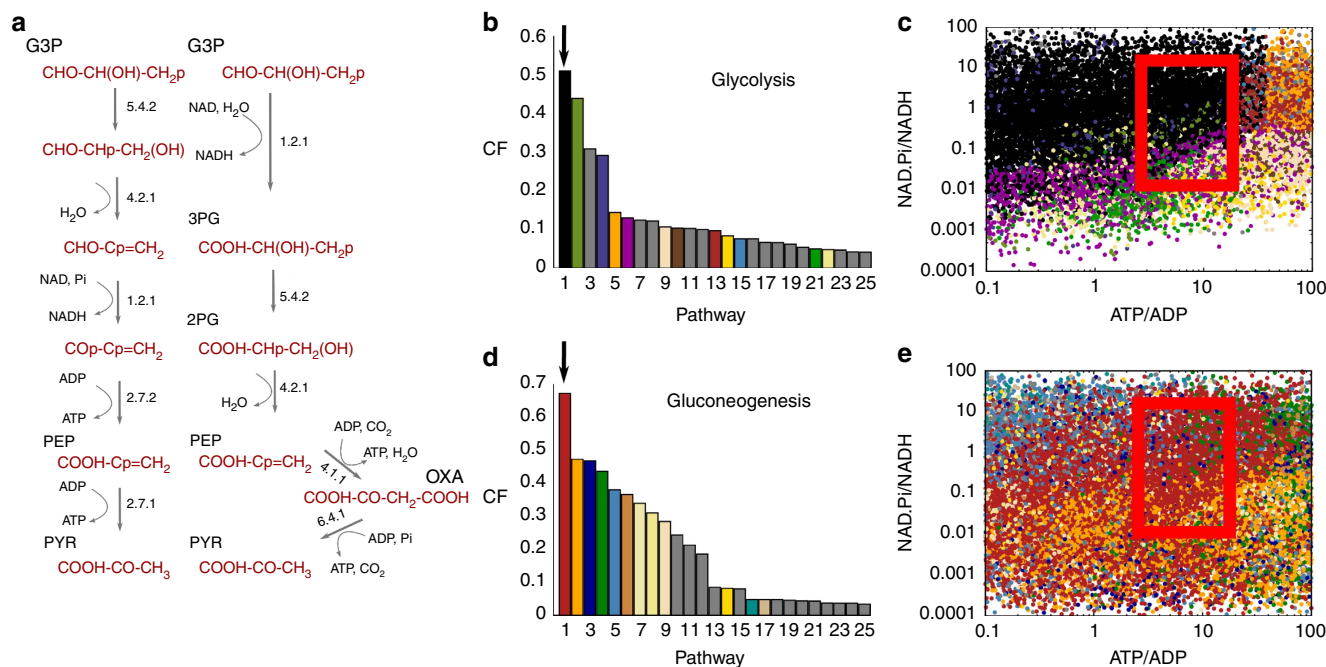
We wish to investigate how many of these alternative trunk pathways are feasible under typical physiological conditions. First, in the glycolytic direction, we demand that candidate pathways produce at least two ATP molecules. This is required so that the full glycolytic pathway produces a net ATP yield, as the two dominant forms of glycolysis, the EMP and ED pathways, consume one ATP molecule per G3P in their upper halves. In addition, motivated by considerations of simplicity<sup>5,10</sup> and cost of enzyme production<sup>8</sup>, we further restrict our analysis to those pathways with a near-minimal number of steps. This results in a set of 1,787 candidate glycolytic pathways of length 4, 5 and 6 that produce at least two ATP molecules, and a set of 6,445 biochemically plausible gluconeogenic pathways of lengths 3–6 (not taking into account thermodynamics).

### The real paths maximize flux under physiological conditions.

We next evaluated the performance of the alternative trunk pathways generated by our network, by comparing their steady-state metabolic flux. For glycolytic pathways, the metabolic flux corresponds to the rate of ATP production, whereas for gluconeogenic pathways, it corresponds to the rate of production of G3P (and ultimately of glucose). This flux depends not only on the total free energy change across a given pathway, but also on the distribution of the individual reaction free energies (see, for example, Fig. 1), which in turn depends on the intracellular environment via the concentrations of the external metabolites. For linear pathways, the flux can be calculated analytically, assuming linear kinetics with diffusion-limited enzymes<sup>7,30,31</sup> (see Methods for details). We also assume that, for each pathway, the individual enzyme concentrations are optimized to maximize pathway flux, for a fixed total amount of enzyme (see Methods for details).

Importantly, when calculating the metabolic flux, we impose constraints on the intermediate internal metabolite concentrations. For the bacterium *Escherichia coli*, metabolite concentrations range from 0.1  $\mu\text{M}$  to 100 mM, with the total intracellular metabolite pool being around 300 mM (ref. 32). We expect that very low metabolite concentrations are undesirable because of vanishing fluxes, whereas very high concentrations are precluded by osmotic considerations. In our initial comparison, we therefore set the flux to zero for a given pathway if any of its intermediate metabolite concentrations falls outside the range of 0.1  $\mu\text{M}$ –100 mM.

Imposing this constraint, we calculated the metabolic flux of all our candidate pathways, across a wide range of intracellular conditions, as defined by the external metabolite concentrations. To explore the performance of our candidate pathways over a wide range of possible intracellular conditions, we randomly selected 10,000 points from the parameter space consisting of the concentrations of the 11 external metabolites, G3P and pyruvate, sampling each parameter logarithmically over several orders of magnitude above and below its typical physiological concentration (see Methods for details). For each point in parameter space, we evaluated the flux of each candidate pathway.



**Figure 2 | The real glycolytic and gluconeogenic trunk pathways represent maximal flux solutions.** (a) Two alternative glycolytic pathways; left-hand pathway uses the same set of reaction classes, defined by EC number, as the real trunk pathway, whereas the right-hand pathway uses a different set. (b–d) Results of pathway comparison. (b,d) Alternative pathways generated in our analysis ordered by their comparative flux (see text for details), averaged across the whole parameter space. The top 25 paths are shown. (b) Glycolytic pathways; the black bar (indicated by the arrow) is the natural pathway. (d) Gluconeogenic paths; the red bar (indicated by the arrow) is the natural pps route. (c,e) Relative pathway performance as a function of the intracellular environment. Each dot represents a randomly sampled point in parameter space; the colour of the dot indicates the candidate pathway, which had the highest flux at that point in parameter space with colours corresponding to b and d. For parameter sampling procedure, see Methods. The axes correspond to two important characteristic measures of a cell, the redox state and energy state, via  $[NAD][Pi]/[NADH]$  and  $[ATP]/[ADP]$ , respectively. The red boxes indicate the typical physiological state of the cell<sup>32,43</sup>:  $[ATP]/[ADP] = 3–20$ ,  $[NAD][Pi]/[NADH] = 0.01–10$  M (based on  $[NAD]/[NADH] = 10–100$  and  $[Pi] = 1–100$  mM). (c) Natural glycolytic pathway (black) tends to perform best under typical physiological conditions. (e) Natural pps gluconeogenic pathway (red) tends to perform best under typical cellular conditions. See Supplementary Data 1 and 2 for pathway structures.

As a simple metric, we first compare the performance of our candidate pathways, averaged over the entire parameter space. To this end, we compute the comparative flux (CF), which is the flux of a given pathway, divided by the maximum flux obtained among all pathways, at a given point in parameter space. Averaging this quantity across the whole parameter space gives a measure of relative performance for each candidate path. Based on this metric, we find that the real glycolytic and gluconeogenic pathways perform remarkably well, compared with the many alternatives (Fig. 2b,d). For glycolysis (Fig. 2b and Supplementary Data 1), the natural trunk pathway (in black, indicated by the arrow) outperforms all the alternative pathways; whereas for gluconeogenesis (Fig. 2d, Supplementary Data 2), one of the real pathways (the pps variant) is also ranked first (in red, indicated by the arrow). These results strongly suggest that the natural trunk pathways carry a high flux compared with alternatives; however, this metric is dependent on the range of parameter space that is sampled.

To overcome this issue and to investigate in more detail, we analysed which of our candidate pathways achieved the highest flux at different points in the parameter space. This allows us to understand how the performance of a given pathway depends on the intracellular environment. Our results show that different candidate pathways perform best in different regions of the parameter space. In particular, pathway performance is very sensitive to the cellular energy state, as measured by the ratio of the ATP and ADP concentrations ( $[ATP]/[ADP]$ ), and redox state, represented by the ratio  $[NAD][Pi]/[NADH]$  (Fig. 2c,e). Focusing on the glycolytic pathways (Fig. 2b) we see that,

remarkably, the natural trunk pathway (black dots) outperforms all the alternatives in the region of parameter space close to that found in living cells (red box). This suggests that the glycolytic trunk pathway represents a maximal flux solution for the conversion of G3P to pyruvate, under typical intracellular conditions. Similarly, for the gluconeogenic pathways (Fig. 2d), the pps pathway found in nature (red dots) outperforms the alternatives under typical physiological conditions (red box).

**Alternative flux calculations.** The flux calculation used above makes the simplifying assumption that all enzymes are perfect catalysts<sup>6,7,30,31</sup> and thus it neglects the important physical feature of substrate saturation. Any attempt at applying a generic choice of kinetic equations to a system of reactions is problematic, and more complicated forms give rise to a greater number of parameters that must be specified. However, one can test the effect that using alternative forms has on the overall outcome, and we find that the above results still hold when two forms of reversible Michaelis–Menten kinetics are assumed (Supplementary Figs 3, 4 and Methods), capturing the behaviour of substrate and product inhibition. Thus, our results are not dependent on the detailed assumptions made in modelling the enzyme kinetics.

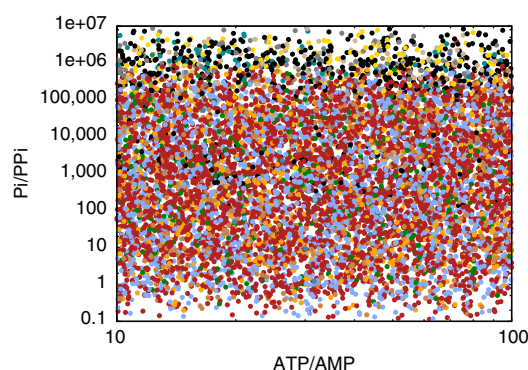
**Constraints on metabolite concentrations are important.** Imposing constraints on the intermediate internal metabolite concentrations is crucial to our analysis. When these constraints are neglected, a very different set of pathways dominate over the parameter space, many of which contain individual reactions with



large  $\Delta_r G$  values, and the real glycolytic pathway falls to number 87 with respect to its CF averaged over parameter space (Supplementary Note 1 and Supplementary Data 3). It is possible, however, that our limits on the concentrations are too severe, as in reality the internal structure of a cell, as well as phenomena such as substrate channelling, can give rise to effective concentrations much larger or smaller than any global limit. Nevertheless, Supplementary Fig. 5 along with Supplementary Data 4 and 5 demonstrates that our results still hold when we repeat our analysis with a more generous allowed concentration range of 1 nM–500 mM.

**Other paths can perform better under different conditions.** We expect the concentrations of the external metabolites to vary across organisms and across different environmental habitats. Hence, an interesting question to ask is could any alternative pathways outperform the real trunk pathways, if the physiological conditions were different? Indeed, we find that for some values of the external metabolite concentrations, other pathways outperform the natural ones (Fig. 2). Of particular interest is the fact that the pyrophosphate concentration is an important factor in our comparison of the gluconeogenic pathways. Figure 3 shows that at low PPi concentrations, the natural ppdk variant of the gluconeogenic pathway becomes more thermodynamically favourable and can outperform the more common pps pathway under typical physiological conditions. In nature, cellular pyrophosphate concentrations can be kept at sub- $\mu$ M levels through the highly efficient inorganic diphosphatase enzyme, EC 3.6.1.1.

**Alternative trunk pathways.** It is interesting to investigate the alternative pathways revealed by our analysis in more detail. Our understanding of even the most well-studied metabolic pathways is incomplete, with new features being discovered in glycolysis<sup>33</sup> and the pentose phosphate pathway<sup>34</sup> in recent years. Hence, it is interesting to consider the alternatives found that most closely resemble the natural pathway, as it is possible that some of these exist in nature, but may be as yet uncharacterized. Figure 4 shows five thermodynamically feasible alternative glycolytic pathways found in our network, which utilize only chemical compounds found in the KEGG COMPOUND database of known metabolites<sup>35</sup> (Supplementary Fig. 6 and Supplementary Data 6). In fact, in three of these pathways (B, D and E), the relevant enzyme for each reaction is known to exist in nature (Supplementary Note 2).

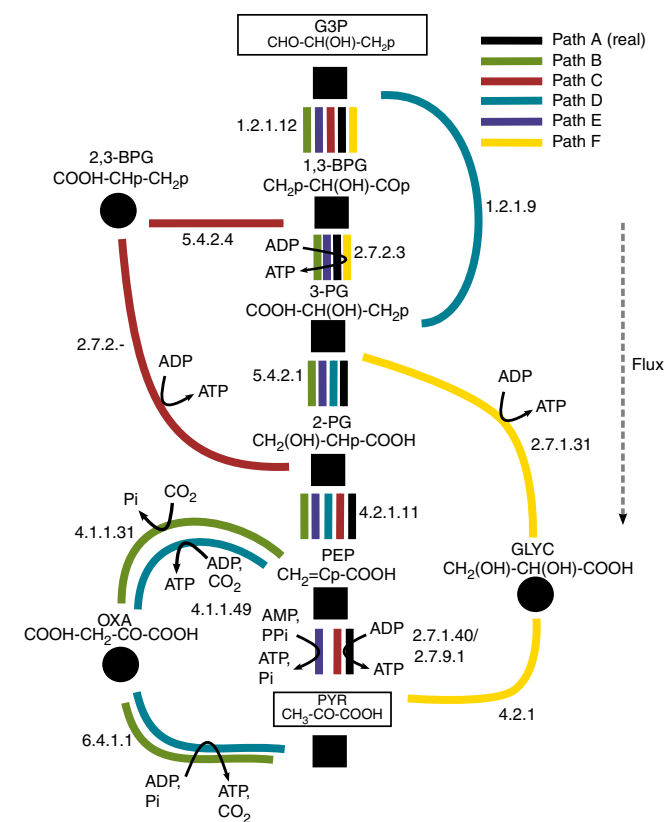


**Figure 3 | The sampled parameter ranges are an important factor in our analysis.** Repeating the comparison of candidate gluconeogenic paths, but now sampling [PPi] from  $10^{-8}$  to  $10^{-2}$  M instead of from  $10^{-4}$  to  $10^{-2}$  shows that the real ppdk pathway (black) becomes favourable compared with the pps pathway (red), when the ratio of orthophosphate to pyrophosphate concentrations is increased. Colours as in Fig. 2.

Path B is a pathway of length 6 and corresponds to the second-highest performing pathway (green) in Fig. 2b. This pathway follows the first four steps of the natural pathway before diverging at PEP to pass through oxaloacetate. This shunt first involves the carboxylation and dephosphorylation of PEP via PEP carboxylase EC 4.1.1.31, followed by a coupled decarboxylation and ATP-generation via pyruvate carboxylase EC 6.4.1.1.

Path C is similar to the real glycolytic trunk path, except that 1,3-bisphosphoglycerate (1,3-BPG) is first isomerized (to 2,3-BPG via EC 5.4.2.4) and then dephosphorylated (to 2-phosphoglycerate (2-PG), with ATP generation), rather than being first dephosphorylated and then isomerized as in the natural pathway. Interestingly, a similar pathway exists in red blood cells, where 2,3-BPG is produced from 1,3-BPG via the Rapoport–Leubering shunt<sup>33</sup>. In red blood cells, however, the 2,3-BPG is hydrolysed to either 3-phosphoglycerate (3-PG) or 2-PG without ATP generation<sup>33</sup>, thus sacrificing one ATP compared with the usual glycolytic pathway. Although no enzyme has been characterized that generates ATP from 2,3-BPG, this reaction does actually appear in the KEGG database, as R02664.

Path D differs from the natural pathway in that it converts G3P directly to 3-PG without the production of ATP, via EC 1.2.1.9. This large exergonic initial step gives path D a favourable thermodynamic profile and results in the two ATP-generating steps being the final two reactions, catalysed by PEP carboxykinase EC 4.1.1.49 and pyruvate carboxylase EC 6.4.1.1. These two enzymes are usually associated with a gluconeogenic role in



**Figure 4 | Thermodynamically feasible alternative glycolytic pathways utilizing compounds found in KEGG<sup>35</sup> (see Supplementary Note 2 for more details).** Reactions are indicated by coloured lines; full EC numbers in black correspond to enzymes present in KEGG, incomplete numbers are not found in KEGG. All of the enzymes for alternative Paths B, D and E are known to exist in nature. GLY, glyceraldehyde; OXA, oxaloacetate; 2,3-BPG = 2,3-bisphosphoglycerate; all others as in Fig. 1.

animals and their feasibility in the glycolytic direction is sensitive to the intracellular pH; it becomes increasingly thermodynamically favourable to produce two ATP molecules from PEP with decreasing pH (see Supplementary Note 2 for details).

Path E is identical to Path A with the exception that the final step generates ATP from AMP and pyrophosphate (via pyruvate phosphate dikinase, EC 2.7.9.1) rather than from ADP and Pi—that is, it is the exact reverse of the natural *ppdk* gluconeogenic pathway. This may result in an effectively greater energetic yield than the natural glycolytic pathway, as an ATP is recovered from an AMP directly, rather than through the action of the adenylate kinase enzyme (EC 2.7.4.3), which involves the consumption of another ATP molecule. This pathway actually corresponds to a glycolytic variant observed in some anaerobic eukaryotes, in which glycolysis is the primary mode of ATP production<sup>36</sup>. Its flux and feasibility will depend on the cellular [AMP] and [PPi] concentrations, which is not the case for the natural glycolytic pathway (which does not involve AMP or PPi).

Finally, Path F represents a pathway of length 4 that produces two ATP molecules. It converts 3-PG to glycerate with the production of ATP, via glycerate 3-kinase, EC 2.7.1.31. Although the final reaction of this pathway is not present in the KEGG database, this putative reaction is chemically rather simple—involving the dehydration of glycerate to enolpyruvate followed by a spontaneous tautomerization to pyruvate—and so it seems plausible that it might be catalysed by the uncharacterized, non-primary action of an existing enzyme.

## Discussion

Despite the huge variety and complexity of life on Earth, the biochemistry of core metabolism is remarkably universal. Our analysis shows that this universality does not arise from an absence of other possibilities. Using a systematic approach, we have identified many alternatives to perhaps the most highly conserved set of metabolic reactions, the glycolytic and gluconeogenic trunk pathways. Our alternative pathways obey the rules of biochemistry, carry positive flux under reasonable intracellular conditions, and satisfy reasonable constraints on metabolite concentrations. Remarkably, of all these alternatives, we find that the trunk pathway observed in nature carries the highest biochemical flux in both the glycolytic and gluconeogenic directions, for parameters that represent typical intracellular physiological conditions. Of the two variants of the prokaryotic gluconeogenic pathway that are found in nature, the *pps* route is the best performer across a wide parameter range, whereas the *ppdk* route is more sensitive to environmental conditions, requiring a low concentration of pyrophosphate (Fig. 3). The fact that our analysis identifies the natural pathways using only flux maximization combined with constraints on intermediate metabolite concentrations suggests that these factors are likely to have been important driving forces in the evolution of metabolism. Importantly, our key results are also independent of the detailed assumptions made in modelling the enzyme kinetics.

Flux maximization is widely recognized as an important concept in the study of metabolism; both from the perspective of glycolysis and more broadly<sup>37</sup>. Previous works addressing this issue for glycolysis<sup>5–7</sup> have arrived at similar conclusions but have remained abstract in nature. For the first time, we attempt to expand on these fundamental and influential studies through the use of a realistic set of generic enzymatic rules along with the inclusion of thermodynamics. Our results add support to the picture that has been developing, and also suggest that evolutionary pressures on metabolic fluxes have to operate within the context of reasonable constraints on metabolite concentrations, and that neglecting these constraints can produce dramatically different outcomes. Our results also

expand on the recent suggestion of Noor *et al.*<sup>10</sup> that central carbon metabolism can be understood as a minimal walk between the set of metabolites essential for growth. We find that the requirement to produce a set of essential biochemical precursors is not sufficient to explain the biochemical structure of the natural trunk pathway, as alternative pathways are possible which produce the essential precursors with the same number of steps (for example, Path D, Fig. 4). Also, many of our alternative pathways produce very similar, but not identical, intermediates to those of the real trunk pathway and it is conceivable that these could be used as alternative precursors. Our results show that flux maximization provides a criterion by which these alternative minimal-length pathways may be distinguished. We also note that this criterion is not affected by ATP yield because alternative pathways producing three or more ATP molecules are never thermodynamically feasible under typical physiological conditions (Supplementary Fig. 7), hence there cannot be any trade-off between the yield and the flux in our model.

Our analysis also reveals alternative trunk pathways that can perform better than the natural one under different physiological conditions. Although some of these alternatives involve compounds and reactions that are not found in biochemical databases, others use enzymes that are known to exist in nature. The latter pathways (Fig. 4) are clearly plausible biochemically and thermodynamically. It could be that some of these pathways do operate glycolytically under some circumstances, and may be revealed by future biochemical investigations (see Supplementary Note 2 for details).

In this study, we have limited our analysis to pathways that start and end at G3P and pyruvate. Relaxing this requirement would certainly lead to many more alternative pathways for the generation of energy and for biosynthesis. Although it is also important to consider other factors, including the need for integration within a wider metabolic network, our analysis suggests that key principles underlying the structure of core metabolism may emerge from simple biochemical, thermodynamic and biophysical considerations.

Finally, we briefly discuss broader evolutionary implications of our work. Assuming that the trunk pathway can be replaced by any of our alternative pathways without affecting the rest of the metabolic network too much, the ATP flux could be used as a proxy for fitness, that is, the ability of the organism to outcompete other, similar to organisms. Endowed with a mutational graph of possible mutations between different pathways (different genotypes), this would create a fitness landscape in which the highest peaks correspond to local maxima of the ATP flux. An interesting question is whether the global maximum (the real pathway) is accessible, that is, can it be reached from anywhere in the landscape? Such fundamental evolutionary questions have been addressed recently using the full set of known biochemical reactions<sup>4</sup>. It would be interesting to repeat this analysis using our data, and to compare the fitness landscape obtained for our model to the landscapes of other evolutionary models<sup>38–41</sup>. This could show to what extent different computer models of artificial life are similar to one another, and perhaps even shed light on the structure of real fitness landscapes.

## Methods

**Chemical compounds and reactions.** We created a list of chemical compounds with 2, 3 or 4 carbon atoms by generating all possible linear combinations of the 20 'building blocks' shown in Supplementary Table 2. Each of the building blocks was composed of a single carbon atom with associated oxygen, hydroxyl, hydrogen, phosphate and/or amino groups. Building blocks were connected together in linear chains by single or double bonds. This procedure created 1,966 linear molecules, 1,477 of which are electrostatically charged in solution, that is, containing at least one carboxyl or phosphate group. These 1,477 molecules are our internal metabolites. Next, for every possible pair of molecules from this list we checked

systematically whether the reactions from Table 1 (see Supplementary Table 3 also for details) could transform one molecule into another, allowing for all possible couplings with the external metabolites. In this way, a network of 7,940 reactions was generated.

**Free energies of compounds and reactions.** For those internal metabolites that are known biochemical species, standard free energies of formation  $\Delta_f G$  were taken from the literature<sup>24</sup>. For other internal metabolites, for which such data does not exist, we employed a variant of the group contribution method<sup>25–28</sup>, which accounts for the fact that molecules exist in solution as an equilibrium mixture of different ionic species. For each such molecule  $g_1 g_2 \dots g_n$ , composed from building blocks  $\{g_i\}$ , we calculated  $\Delta_f G$  using

$$\Delta_f G = E_0 + \sum_j E_1(g_j) + \sum_{\langle j,k \rangle} E_2(g_j, g_k), \quad (1)$$

where  $E_0$  is a constant,  $E_1(g_j)$  is the contribution of group  $g_j$  and  $E_2(g_j, g_k)$  is a small correction due to neighbouring group–group interactions. The values of  $E_0$ , the vector  $E_1$  and matrix  $E_2$  are determined by performing a least-squares fit to a training set of molecules with known  $\Delta_f G$ s that correspond most closely to the linear CHOPN molecules of our network (see Supplementary Methods for details).

**Flux calculation.** We used the method mentioned in ref. 7 to calculate the flux carried by a linear pathway. This method assumes that the flux through reaction  $i$  is given by<sup>7,30</sup>

$$v_i = \frac{k_d [E_i] ([S_{i-1}] q_i - [S_i])}{1 + q_i}, \quad (2)$$

where  $k_d$  is the diffusion-controlled rate constant,  $[E_i]$  is the enzyme concentration,  $[S_{i-1}]$  and  $[S_i]$  represent substrate and product concentrations and  $q_i$  is the thermodynamic constant. This expression assumes that the enzyme acts as a perfect catalyst, and is used to derive an expression for pathway flux and metabolite concentrations (see Supplementary Methods for details). We then use Powell's method<sup>42</sup> to find the set of enzyme concentrations that maximize the flux subject to the constraints that (i) all steady-state intermediate concentrations are within the prescribed range and (ii) the total enzyme concentration is fixed. We also repeat our analysis using reversible Michaelis–Menten kinetics, see Supplementary Methods for details of calculation.

**Sampling the parameter space.** We randomly selected 10,000 points from the parameter space corresponding to the concentrations of 11 external metabolites and the G3P and pyruvate concentrations. Each parameter was sampled logarithmically over a range covering several orders of magnitude above and below its typical physiological concentrations (see Supplementary Table 4 for details). For each of these points, we calculated the optimized flux  $J_i$  of each candidate pathway as detailed above, and computed the CF of path  $i$  as  $CF_i = J_i / \max\{J_k\}$ , by dividing its flux by the highest flux obtained across all pathways at the given point in parameter space.

**Robustness of our results to small free energy changes.** Using the group contribution method, the typical error in our calculation of the free energy of formation  $\Delta_f G$  for a given molecule is a few  $\text{kJ mol}^{-1}$  (see Supplementary Methods for details). To check the robustness of our results to such errors, our entire analysis was repeated using  $\Delta_f G$  values computed using different sets of training molecules, consisting of 80% of the molecules from the original training set, chosen at random. The qualitative results using such networks were identical from those obtained from the full set of training compounds. For example, the top 25 glycolytic pathways obtained from the reduced set contained 23 out of the 25 pathways from the original analysis.

## References

- Romano, A. & Conway, T. Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* **147**, 448–455 (1996).
- Smith, E. & Morowitz, H. J. Universality in intermediary metabolism. *Proc. Natl Acad. Sci. USA* **101**, 13168–13173 (2004).
- Pal, C. *et al.* Chance and necessity in the evolution of minimal metabolic networks. *Nature* **440**, 667–670 (2006).
- Barve, A., Hosseini, S.-R., Martin, O. C. & Wagner, A. Historical contingency and the gradual evolution of metabolic properties in central carbon and genome-scale metabolisms. *BMC Syst. Biol.* **8**, doi:10.1186/1752-0509-8-48 (2014).
- Meléndez-Hevia, E. & Isidoro, A. The game of the pentose phosphate cycle. *J. Theor. Biol.* **117**, 251–263 (1985).
- Ebenhöh, O. & Heinrich, R. Evolutionary optimization of metabolic pathways. theoretical reconstruction of the stoichiometry of ATP and NADH producing systems. *B. Math. Biol.* **63**, 21–55 (2001).
- Heinrich, R., Montero, F., Klipp, E., Waddell, T. G. & Meléndez-Hevia, E. Theoretical approaches to the evolutionary optimization of glycolysis: thermodynamic and kinetic constraints. *Eur. J. Biochem.* **243**, 191–201 (1997).
- Flamholz, A., Noor, E., Bar-Even, A., Liebermeister, W. & Milo, R. Glycolytic strategy as a tradeoff between energy yield and protein cost. *Proc. Natl Acad. Sci. USA* **110**, 10039–10044 (2013).
- Barve, A. & Wagner, A. A latent capacity for evolutionary innovation through exaptation in metabolic system. *Nature* **500**, 203–206 (2013).
- Noor, E., Eden, E., Milo, R. & Alon, U. Central carbon metabolism as a minimal biochemical walk between precursors for biomass and energy. *Mol. Cell* **39**, 809–820 (2010).
- Samal, A., Wagner, A. & Martin, O. C. Environmental versatility promotes modularity in genome-scale metabolic networks. *BMC Syst. Biol.* **5**, 135 (2011).
- Bar-Even, A., Noor, E., Lewis, N. E. & Milo, R. Design and analysis of synthetic carbon fixation pathways. *Proc. Natl Acad. Sci. USA* **107**, 8889–8894 (2010).
- Handorf, T., Ebenhöh, O. & Heinrich, R. Expanding metabolic networks: Scopes of compounds, robustness, and evolution. *J. Mol. Evol.* **61**, 498–512 (2005).
- Meléndez-Hevia, E., Waddell, T. G., Heinrich, R. & Montero, F. Theoretical approaches to the evolutionary optimization of glycolysis: chemical analysis. *Eur. J. Biochem.* **244**, 527–543 (1997).
- Mittenthal, J. E., Yuan, A., Clarke, B. & Scheeline, A. Designing metabolism: Alternative connectivities for the pentose phosphate pathway. *Bull. Math. Biol.* **60**, 815–856 (1998).
- Hatzimanikatis, V. *et al.* Exploring the diversity of complex metabolic networks. *Bioinformatics* **21**, 1603–1609 (2005).
- Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate. *Biotechnol. Bioeng.* **106**, 462–473 (2010).
- Bar-Even, A., Flamholz, A., Noor, E. & Milo, R. Rethinking glycolysis: on the biochemical logic of metabolic pathways. *Nat. Chem. Biol.* **8**, 509–517 (2012).
- Ronimus, R. S. & Morgan, H. W. Distribution and phylogenies of enzymes of the Embden-Meyerhof-Parnas pathway from archaea and hyperthermophilic bacteria support a gluconeogenic origin of metabolism. *Archaea* **1**, 199–221 (2002).
- Verhees, C. H. *et al.* The unique features of glycolytic pathways in Archaea. *Biochem. J.* **375**, 231–246 (2003).
- Morowitz, H. J., Kostelnik, J. D., Yang, J. & Cody, G. D. The origin of intermediary metabolism. *Proc. Natl Acad. Sci. USA* **97**, 7704–7708 (2000).
- Morowitz, H. J. A theory of biochemical organization, metabolic pathways, and evolution. *Complexity* **4**, 39–53 (1999).
- Srinivasan, V. & Morowitz, H. J. Analysis of the intermediary metabolism of a reductive chemoautotroph. *Biol. Bull.* **217**, 222–232 (2009).
- Alberty, R. A. Biochemical thermodynamics: applications of mathematica. *Methods Biochem. Anal.* **48**, 1–458 (2006).
- Mavrovouniotis, M. L. Estimation of standard Gibbs energy changes of biotransformations. *J. Biol. Chem.* **266**, 14440–14445 (1991).
- Mavrovouniotis, M. L. Group contributions for estimating standard Gibbs energies of formation of biochemical compounds in aqueous solution. *Biotechnol. Bioeng.* **36**, 1070–1082 (1990).
- Jankowski, M. D., Henry, C. S., Broadbelt, L. J. & Hatzimanikatis, V. Group contribution method for thermodynamic analysis of complex metabolic networks. *Biophys. J.* **95**, 1487–1499 (2008).
- Noor, E. *et al.* An integrated open framework for thermodynamics of reactions that combines accuracy and coverage. *Bioinformatics* **28**, 2037–2044 (2012).
- Webb, E. *Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes* 6 edn (Academic Press, 1992).
- Albery, J. W. & Knowles, J. R. Evolution of enzyme function and the development of catalytic efficiency. *Biochemistry* **15**, 5631–5640 (1976).
- Heinrich, R. & Hoffmann, E. Kinetic parameters of enzymatic reactions in states of maximal activity; an evolutionary approach. *J. Theor. Biol.* **151**, 249–283 (1991).
- Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat. Chem. Biol.* **5**, 593–599 (2009).
- Cho, J., King, J. S., Qian, X., Harwood, A. J. & Shears, S. B. Dephosphorylation of 2,3-bisphosphoglycerate by MIPP expands the regulatory capacity of the Rapoport-Luebering glycolytic shunt. *Proc. Natl Acad. Sci. USA* **105**, 5998–6003 (2008).
- Stincone, A. *et al.* The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. *Biol. Rev.* **90**, 927–963 (2015).
- Kyoto encyclopedia of genes and genomes. <http://www.genome.jp/kegg/> (2015).
- Feng, X.-M., Cao, L.-J., Adam, R. D., Zhang, X.-C. & Lu, S.-Q. The catalyzing role of PPK in *Giardia lamblia*. *Biochem. Biophys. Res. Commun.* **367**, 394–398 (2008).
- Varma, A. & Palsson, B. O. Metabolic flux balancing - basic concepts, scientific and practical use. *Nature Biotech.* **12**, 994–998 (1994).
- Hintze, A. & Adami, C. Evolution of complex modular biological networks. *PLoS Comput. Biol.* **4**, e23 (2008).
- Östman, B. & Adami, C. In *Recent Advances in the Theory and Application of Fitness Landscapes*, vol. 6 of *Emergence, Complexity and Computation*. (eds Richter, H. & Engelbrecht, A.) 509–526 (Springer Berlin Heidelberg, 2014).
- Lenski, R. E., Ofria, C., Pennock, R. T. & Adami, C. The evolutionary origin of complex features. *Nature* **423**, 139–144 (2003).



41. Ullrich, A., Rohrschneider, M., Scheuermann, G., Stadler, P. F. & Flamm, C. In silico evolution of early metabolism. *Artif. Life* **17**, 87–108 (2011).
42. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes: The Art of Scientific Computing* 3rd edn (Cambridge Univ., 2007).
43. Berg, J., Hung, Y. P. & Yellen, G. A genetically encoded fluorescent reporter of ATP:ADP ratio. *Nat. Methods* **6**, 161–166 (2009).

## Acknowledgements

B.W. and R.J.A. contributed equally to this work. We thank Patrick Warren, Vincent Danos and Oliver Ebenhöf for insightful discussions. This work was funded by EPSRC under Grant No. EP/J007404/1 and by HFSP under grant RGY0081/2012. S.J.C. was supported by a Carnegie/Caledonian Scholarship from the Carnegie Trust for Scotland, B.W. was supported by a Leverhulme Trust Early Career Fellowship and by a Royal Society of Edinburgh Personal Research Fellowship, and R.J.A. was supported by a Royal Society University Research Fellowship.

## Author contributions

S.C., B.W. and R.J.A. designed the study. S.C. and B.W. generated the network, B.W. performed the thermodynamic calculations, S.C. conducted the simulations and analysis. All authors contributed to writing the manuscript.

## Additional information

**Supplementary Information** accompanies this paper at <http://www.nature.com/naturecommunications>

**Competing financial interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**How to cite this article:** Court, S. J. *et al.* Lower glycolysis carries a higher flux than any biochemically possible alternative. *Nat. Commun.* 6:8427 doi: 10.1038/ncomms9427 (2015).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>